REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggesstions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any oenalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
27-05-2015	Final Report	1-Jul-2010 - 31-Aug-2014

4. TITLE AND SUBTITLE
Final Report: Predictive Anomaly Management for Resilient Virtualized Computing Infrastructures

5a. CONTRACT NUMBER
W911NF-10-1-0273

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
Xiaohui Gu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
North Carolina State University
2701 Sullivan Drive

Raleigh, NC 27695 -7514

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
56351-CS.18

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
Large-scale distributed virtualized computing infrastructures have become important platforms for many critical real-world systems such as cloud computing, big data processing, and intelligence analysis. However, due to its inherent complexity and sharing nature, virtualized computing infrastructures are inevitably prone to various system anomalies caused by software bugs, hardware failures, and resource contentions. The situation exacerbates if the system is also exposed to malicious attacks. Moreover, although some anomaly symptoms such as machine crash are easy to detect, many other anomalies (e.g., performance degradation, processing bottlenecks, memory

15. SUBJECT TERMS
Anomaly Prediction, Anomaly Diagnosis, Virtualized Computing Infrastructure

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Xiaohui (Helen) Gu
UU	UU	UU	UU		19b. TELEPHONE NUMBER 919-515-7045

Report Title

Final Report: Predictive Anomaly Management for Resilient Virtualized Computing Infrastructures

ABSTRACT

Large-scale distributed virtualized computing infrastructures have become important platforms for many critical real-world systems such as cloud computing, big data processing, and intelligence analysis. However, due to its inherent complexity and sharing nature, virtualized computing infrastructures are inevitably prone to various system anomalies caused by software bugs, hardware failures, and resource contentions. The situation exacerbates if the system is also exposed to malicious attacks. Moreover, although some anomaly symptoms such as machine crash are easy to detect, many other anomalies (e.g., performance degradation, processing bottlenecks, memory leak bugs) are hard to detect and diagnosis, which often have latent impact to the system. In this project, we have explored various online system anomaly prediction and cause inference schemes using unsupervised machine learning methods. We tested our algorithms using extensive real system experiments and our results show that we can achieve high fidelity anomaly prediction (i.e., >95% true positive rate with <1% false positive rate) with low overhead (<1% CPU load).

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper
05/18/2015 15.00	Hiep Nguyen , Daniel J. Dean, Kamal Kc, Xiaohui Gu. Insight: In-situ Online Service Failure Path Inference in Production Computing Infrastructures, USENIX Annual Technical Conference (ATC). 19-JUN-14, . : ,
05/18/2015 16.00	Daniel J. Dean, Hiep Nguyen, Peipei Wang, Xiaohui Gu. PerfCompass: Toward Runtime Performance Anomaly Fault Localization for Infrastructure-as-a-Service Clouds, USENIX Workshop on Hot Topics in Cloud Computing . 17-JUN-14, . : ,
TOTAL:	**2**

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper
05/18/2015	14.00 Tsung-Hsuan Ho, Daniel Dean, Xiaohui Gu, William Enck. PREC: Practical Root Exploit Containment for Android Devices, ACM Conference on Data and Application Security and Privacy (CODASPY) . 03-MAR-14, . : ,
05/18/2015	17.00 Daniel J. Dean , Hiep Nguyen, Xiaohui Gu , Hui Zhang , Junghwan Rhee , Nipun Arora , Geoff Jiang. PerfScope: Practical Online Server Performance Bug Inference in Production Cloud Computing Infrastructures, ACM Symposium on Cloud Computing (SOCC). 03-NOV-14, . : ,
07/03/2013	11.00 Hiep Nguyen, Zhiming Shen, Yongmin Tan, Xiaohui Gu. FChain: Toward Black-box Online Fault Localization for Cloud Systems, IEEE International Conference on Distributed Computing Systems (ICDCS). 09-JUL-13, . : ,
07/03/2013	12.00 Hiep Nguyen, Zhiming Shen, Xiaohui Gu, Sethuraman Subbiah, John Wilkes. AGILE: elastic distributed resource scaling for Infrastructure-as-a-Service, USENIX International Conference on Autonomic Computing (ICAC). 26-JUN-13, . : ,
08/06/2012	7.00 Zhiming Shen, Sethuraman Subbiah, Xiaohui Gu, John Wilkes. CloudScale: Elastic Resource Scaling for Multi-Tenant Cloud Systems, CloudScale: Elastic Resource Scaling for Multi-Tenant Cloud Systems. 14-OCT-11, . : ,
08/06/2012	8.00 Yongmin Tan, Hiep Nguyen, Zhiming Shen, Xiaohui Gu, Chitra Venkatramani, Deepak Rajan. PREPARE: Predictive Performance Anomaly Prevention for Virtualized Cloud Systems, Proc. of International Conference on Distributed Computing Systems (ICDCS). 18-JUN-12, . : ,
08/14/2011	1.00 Ying Zhao, Yongmin Tan, Zhenhuan Gong, Xiaohui Gu, Mike Wamboldt. Self-Correlating Predictive Information Tracking for Large-Scale Production Systems, ?IEEE International Conference on Autonomic Computing and Communications (ICAC). 15-JUN-09, . : ,
08/14/2011	2.00 Yongmin Tan, Vinay Venkatesh, Xiaohui Gu. OLIC: OnLine Information Compression for Scalable Distributed System Monitoring, ACM/IEEE International Workshop on Quality of Service (IWQoS). 06-JUN-11, . : ,
08/14/2011	3.00 Yongmin Tan, Xiaohui Gu, Haixun Wang. Adaptive Runtime Anomaly Prediction for Dynamic Hosting Infrastructures, ACM Symposium on Principles of Distributed Computing (PODC). 25-JUL-10, . : ,
08/14/2011	4.00 Yongmin Tan, Xiaohui Gu. On Predictability of System Anomalies in Real World, IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS). 17-AUG-10, . : ,
08/14/2011	5.00 Kamal Kc, Xiaohui Gu. ELT: Efficient Log-based Troubleshooting System for Cloud Computing Infrastructures, IEEE International Symposium on Reliable Distributed Systems (SRDS). 05-OCT-11, . : ,
08/14/2011	6.00 Hiep Nguyen, Yongmin Tan, Xiaohui Gu. Propagation-aware Anomaly Localization for Cloud Hosted Distributed Applications, ACM Workshop on Managing Large-Scale Systems via the Analysis of System Logs and the Application of Machine Learning Techniques (SLAML) in conjunction with SOSP. 23-OCT-11, . : ,

08/14/2012 9.00 Hiep Nguyen, Daniel Dean, Xiaohui Gu. UBL: Unsupervised Behavior Learning for Predicting Performance Anomalies in Virtualized Cloud System,
Proc. of International Conference on Autonomic Computing (ICAC). 17-SEP-12, . : ,

TOTAL: **13**

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received Paper

TOTAL:

Number of Manuscripts:

Books

Received Book

07/03/2013 13.00 Daniel J. Dean, Yongmin Tan, Xiaohui Gu , Ting Yu, Juan Du. Scalable Distributed Service Integrity Attestation for Software-as-a-Service Clouds", IEEE Transactions on Parallel and Distributed Systems : IEEE computer soceity, (03 2013)

08/14/2012 10.00 Yongmin Tan, Vinay Venkatesh, Xiaohui Gu. Resilient Self-Compressive Monitoring for Large-Scale Hosting Infrastructures, IEEE Transactions on Parallel and Distributed Systems: IEEE Transactions, (05 2012)

TOTAL: **2**

Received	Book Chapter

TOTAL:

Patents Submitted

Unsupervised Behavior Learning System and Method for Predicting Performance Anomalies in Distributed Computing Infrastructures

Patents Awarded

Awards

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Daniel Dean	1.00	
Hiep Nguyen	1.00	
Anwesha Das	1.00	
Peipei Wang	1.00	
FTE Equivalent:	**4.00**	
Total Number:	**4**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Xiaohui Gu	0.18	
FTE Equivalent:	**0.18**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED

FTE Equivalent:
Total Number:

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period:	0.00
The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:.....	0.00
The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:.....	0.00
Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):.....	0.00
Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:.....	0.00
The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense	0.00
The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:	0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME
Hiep Nguyen
Daniel Dean

Total Number: **2**

Names of other research staff

NAME	PERCENT_SUPPORTED

FTE Equivalent:
Total Number:

Sub Contractors (DD882)

Inventions (DD882)

5 **Unsupervised Behavior Learning System and Method for Predicting Performance Anomalies in Distributed Computing In**

Patent Filed in US? (5d-1) Y

Patent Filed in Foreign Countries? (5d-2) N

Was the assignment forwarded to the contracting officer? (5e) N

Foreign Countries of application (5g-2):

5a: Daniel Dean

5f-1a: North Carolina State University

5f-c: 890 Oval Drive

Raleigh NC 27695

5a: Xiaohui Gu

5f-1a: North Carolina State University

5f-c: 890 Oval Drive

Raleigh NC 27695

Scientific Progress

Objective

Larg-scale distributed computing infrastructures have become important platforms for many critical real-world systems such as cloud computing, big data processing, and intelligence analysis. However, due to its inherent complexity and sharing nature, shared computing infrastructures are inevitably prone to various system anomalies caused by software bugs, hardware failures, and resource contentions. The situation exacerbates if the system is also exposed to malicious attacks. Moreover, although some anomaly symptoms such as machine crash are easy to detect, many other anomalies (e.g., performance degradation, processing bottlenecks, memory leak bugs) are hard to detect and diagnosis, which often have latent impact to the system. The objective of this project is to develop automatic 24x7 anomaly management to enhance the resilience of large-scale shared computing infrastructures.

Approach

In this project, we propose to develop a new predictive anomaly management approach that can raise advance anomaly alerts to trigger just-in-time anomaly diagnosis while the system approaches the anomaly state, and perform informed anomaly correction based on the runtime diagnosis results before the system is seriously affected by the anomaly. Thus, our approach can effectively alleviate the impact of anomalies without incurring prohibitive cost to the infrastructure. We focus on developing novel techniques for predicting, diagnosing, and correcting latent anomalies in shared computing infrastructures. The latent anomalies (e.g., performance degradation, resource hotspots, memory leak bugs) often do not have salient symptoms at the beginning, which make it hard to detect by human being. Those latent anomalies are often difficult to diagnose since their symptoms are often correlated with many reasons. However, it is highly important to detect and correct those latent anomalies since they often have prolonged impact to the system. We test our techniques on not only controllable virtual computing systems running in our lab but also on production-level infrastructures such as virtual computing lab (VCL) at NCSU and real world computing infrastructure data provided by our industrial partners at Google and IBM. We also develop metrics and models to evaluate the predictability of a wide range of system anomalies so as to build taxonomy of predictable system anomalies. We also develop new prediction and containment techniques to prevent root exploit attacks on edge-devices such as smart phones, which are the most serious attacks among all the security attacks and are hard to prevent using exiting techniques.

Scientific Barriers

Statistical learning and detailed data analysis have recently been shown to be promising for automatic system status analysis. Our work leverages statistical learning and signal processing techniques to achieve online anomaly prediction. The major challenge includes how to achieve high prediction accuracy under dynamic computing environments and raise early enough alerts before anomaly happens. We have developed various online anomaly prediction techniques to achieve this goal. We developed prediction algorithms using both supervised and unsupervised learning techniques. The unsupervised learning approach allows us to achieve online anomaly prediction without requiring anomaly training data. Thus, our techniques can predict both previously known and unknown anomalies. We also developed context-aware anomaly prediction techniques that can achieve much higher prediction accuracy for dynamic systems than previous schemes. We recently extend our prediction algorithm that can consider not only system-level metrics (e.g., CPU, memory, disk usage) but also system calls. By analyzing system calls, our prediction algorithm can successfully predict all the existing root exploit attacks on the Android smart phones.

Prediction enables us to trigger timely preventions (e.g., migration, resource scaling, inserting delays in system calls) before the user perceives serious impact from the anomaly. We developed various online anomaly prevention techniques using live virtual machine (VM) migrations and elastic resource scaling. Our prediction system not only can raise advance alerts but also provide root cause inference to identify what might be the root cause of the system anomaly (e.g., CPU hog, memory leak, disk contention). We can then invoke proper prevention actions accordingly. Since prediction might raise false alarms, we also develop validation schemes to reverse incorrect preventions.

Prediction also enables us to perform in-situ anomaly diagnosis that can identify anomaly root causes onsite. The advantage is that we don’t need to reproduce the anomaly-inducing environments, which are often extremely difficult. We are developing onsite anomaly path inference and various root cause localization techniques. We can first localize the faulty components among many distributed system components. We then localize root cause functions using system call analysis. The basic idea is to learn the system call sequence patterns produced by different functions using frequent episode mining and then use those system call sequence patterns as signatures to identify root cause functions. The advantage of our approach is that we don’t require source code or any high-overhead online system instrumentations. We also develop onsite failure path inference without requiring source code.

Virtual machines provide opportunities for us to monitor and control various applications running inside the computing infrastructure. Our work leverages virtual machines to perform out-of-box monitoring and control. One challenge we have

addressed in this project is to achieve scalable runtime monitoring, which can continuously track different virtual machine (VM) execution data (e.g., performance counters, resource metrics, system calls, inter-component invocations) to provide comprehensive knowledge for anomaly prediction and diagnosis. We developed adaptive sampling and online compression techniques to achieve light-weight monitoring.

Significance

The proposed research fundamentally advances knowledge and understanding in the interdisciplinary field of applying machine learning and dynamic system analysis to improve the resilience of complex computing infrastructures. Enhancing the resilience of large-scale computing infrastructures, which is well recognized by ARO as one of its key computing challenges in future battle spaces. As more and more critical Army missions depend on IT infrastructure, it has become imperative to guarantee continuous system operation despite software/hardware failures and malicious attacks. As rapid advances in computing hardware have led to dramatic improvement in computer performance, the issues of reliability, availability, and manageability are becoming the nominating bottlenecks in IT infrastructure maintenance. The proposed research advances existing science and technology through novel techniques in support of self-evolving system modeling, online anomaly prediction, onsite anomaly diagnosis, and anomaly preventions for large-scale distributed computing infrastructure. The proposed research explores new approaches with novel applications of machine learning, speculative execution, and dynamic system analysis on system profiling, anomaly prediction and diagnosis, and development of new scalable techniques and tools to achieve resilient distributed computing systems. We will develop and make available implemented techniques and collected data, which will let other researchers and practitioners build on our results.
Accomplishments
(feel free to use a bulleted list here)

Publications:
- "PerfScope: Practical Online Server Performance Bug Inference in Production Cloud Computing Infrastructures", Daniel Dean, Hiep Nguyen, Xiaohui Gu, Hui Zhang, Junghwan Rhee, Nipun Arora, Geoff Jiang, Proc. of ACM Symposium on Cloud Computing (SOCC), Seattle, WA, November, 2014. (acceptance rate: 29/119 = 24%)
- "PerfCompass: Toward Runtime Performance Anomaly Fault Localization for Infrastructure-as-a-Service Clouds", Daniel Dean, Hiep Nguyen, Peipei Wang, Xiaohui Gu,
Proc. of USENIX Workshop on Hot Topics in Cloud Computing (HotCloud), Philadelphia, PA, June, 2014. (acceptance rate: 22/72 = 30.5%)
- "Insight: In-situ Online Service Failure Path Inference in Production Computing Infrastructures", Hiep Nguyen, Daniel J. Dean, Kamal Kc, Xiaohui Gu
Proc. of USENIX Annual Technical Conference (USENIX ATC), Philadelphia, PA, June, 2014. (acceptance rate: 36/241 = 14.9%)
- "PREC: Practical Root Exploit Containment for Android Devices", Tsung-Hsuan Ho, Daniel Dean, Xiaohui Gu, William Enck, Proc. of the ACM Conference on Data and Application Security and Privacy (CODASPY), San Antonio, TX, March, 2014. (full paper, acceptance rate: 16%)
- "AGILE: elastic distributed resource scaling for Infrastructure-as-a-Service", Hiep Nguyen, Zhiming Shen, Xiaohui Gu, Sethuraman Subbiah, John Wilkes, Proc. of USENIX International Conference on Autonomic Computing (ICAC), San Jose, CA, June, 2013. (full paper, acceptance rate: 16/73 = 21%)
- "FChain: Toward Black-box Online Fault Localization for Cloud Systems", Hiep Nguyen, Zhiming Shen, Yongmin Tan, Xiaohui Gu, Proc. of IEEE International Conference on Distributed Computing Systems (ICDCS), Philadelphia, PA, July, 2013. (acceptance rate: 61/464 = 13%)
- "Scalable Distributed Service Integrity Attestation for Software-as-a-Service Clouds", Juan Du, Daniel Dean, Yongmin Tan, Xiaohui Gu, Ting Yu, IEEE Transactions on Parallel and Distributed Systems (TPDS), 2013.
- "UBL: Unsupervised Behavior Learning for Predicting Performance Anomalies in Virtualized Cloud Systems", Daniel Dean, Hiep Nguyen, Xiaohui Gu, Proc. of International Conference on Autonomic Computing (ICAC), San Jose, CA, September, 2012. (acceptance rate: 24%)
- "PREPARE: Predictive Performance Anomaly Prevention for Virtualized Cloud Systems", Yongmin Tan, Hiep Nguyen, Zhiming Shen, Xiaohui Gu, Chitra Venkatramani, Deepak Rajan, Proc. of International Conference on Distributed Computing Systems (ICDCS), Macau, China, June, 2012 (acceptance rate: 71/515=13.8%, best paper award).
- "Resilient Self-Compressive Monitoring for Large-Scale Hosting Infrastructures", Yongmin Tan, Vinay Venkatesh, Xiaohui Gu, IEEE Transactions on Parallel and Distributed Systems (TPDS), 2012.
- "Propagation-aware Anomaly Localization for Cloud Hosted Distributed Applications",Hiep Nguyen and Yongmin Tan and Xiaohui Gu, Proc. of ACM Workshop on Managing Large-Scale Systems via the Analysis of System Logs and the Application of Machine Learning Techniques (SLAML) in conjunction with SOSP, Cascais, Portugal, October, 2011.
- "ELT: Efficient Log-based Troubleshooting System for Cloud Computing Infrastructures", Kamal Kc, Xiaohui Gu, Proc. of IEEE International Symposium on Reliable Distributed Systems (SRDS), Madrid, Spain, October, 2011.
- "OLIC: OnLine Information Compression for Scalable Distributed System Monitoring", Yongmin Tan, Vinay Venkatesh,

Xiaohui Gu, Proc. of ACM/IEEE International Workshop on Quality of Service (IWQoS), San Jose, CA, June, 2011.
- "Adaptive Runtime Anomaly Prediction for Dynamic Hosting Infrastructures", Yongmin Tan, Xiaohui Gu, Haixun Wang,ACM Symposium on Principles of Distributed Computing (PODC), Zurich, Switzerland, July, 2010. (Acceptance rate: 21%)
- "PRESS: PRedictive Elastic ReSource Scaling for Cloud Systems", Zhenhuan Gong, Xiaohui Gu, John Wilkes, IEEE International Conference on Network and Services Management (CNSM), Niagara Falls, Canada, October, 2010.(acceptance rate: 27/176 = 15%, Best Paper Award)
- "On Predictability of System Anomalies in Real World", Yongmin Tan, Xiaohui Gu,Proc. of IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS), Miami Beach, Florida, August, 2010. (Acceptance rate: 29%)
- "Self-Correlating Predictive Information Tracking for Large-Scale Production Systems",Ying Zhao, Yongmin Tan, Zhenhuan Gong, Xiaohui Gu, Mike Wamboldt, IEEE International Conference on Autonomic Computing and Communications (ICAC), Barcelona, Spain, June, 2009. (Acceptance rate: 15.6%)

Awards:
- Best paper awards, IEEE ICDCS, 1 out of 530 submissions, 2012.
- Best paper awards, IEEE CNSM, 1 out of 176 submissions, 2010.

Media coverage:
- featured highlight on NSF's official news site, Science 360,
- Communications of ACM,
- e! science news,
- WRAL techwire,
- ScienceDaily, etc.

Collaborations and Leveraged Funding

We have been collaborating with VCL administrators to apply our techniques on the VCL infrastructure. Most of our tools have been tested on the VCL. We have been working with researchers at IBM and Google during this project. Our current leveraged funding include:

- "CAREER: Enabling Robust Virtualized Hosting Infrastructures via Coordinated Learning, Recovery, and Diagnosis", NSF, $450K, 1/1/2012-12/31/2016, Sole PI.
- "Deepening the Understanding of Least Privilege Through Automatic Partitioning of Hybrid Programs", NSA Science of Security Lablet, $522K, 1/1/2012-12/31/2014, Co-PI, PI: William Enck.
- "Online Performance Anomaly Diagnosis for Cloud Computing Infrastructures", IBM Faculty Award, $15K, 9/1/2011-8/31/2012, Sole PI.
- "CSR:Small: Online System Anomaly Prediction and Diagnosis for Large-Scale Hosting Infrastructures",NSF, $405,000, 08/15/2009 – 08/14/2012,Sole PI.

Conclusions

We have successfully integrated our online anomaly prediction, anomaly root cause inference, and anomaly prevention components into a complete automatic anomaly prevention framework. Our system can automatically steer the system away from anomalies caused by various software bugs, resource contentions, or malicious attacks.

Technology Transfer

NCSU filed a patent application on our unsupervised anomaly prediction scheme and Google has purchased an evaluation license for our software.

Technology Transfer

NCSU filed a patent on our unsupervised behavior learning (UBL) technology. Google has entered an evaluation agreement with NCSU for licensing UBL. One startup is underway to commercialize our anomaly prediction and diagnosis techniques.

Predictive Anomaly Management for Resilient Computing Infrastructures
Proposal Number (56351-CS)
Professor Xiaohui Helen Gu, North Carolina State University

Objective

Larg-scale distributed computing infrastructures have become important platforms for many critical real-world systems such as cloud computing, big data processing, and intelligence analysis. However, due to its inherent complexity and sharing nature, shared computing infrastructures are inevitably prone to various system anomalies caused by software bugs, hardware failures, and resource contentions. The situation exacerbates if the system is also exposed to malicious attacks. Moreover, although some anomaly symptoms such as machine crash are easy to detect, many other anomalies (e.g., performance degradation, processing bottlenecks, memory leak bugs) are hard to detect and diagnosis, which often have latent impact to the system. The objective of this project is to develop automatic 24x7 anomaly management to enhance the resilience of large-scale shared computing infrastructures.

Approach

In this project, we propose to develop a new *predictive* anomaly management approach that can raise advance anomaly alerts to trigger just-in-time anomaly diagnosis while the system approaches the anomaly state, and perform informed anomaly correction based on the runtime diagnosis results before the system is seriously affected by the anomaly. Thus, our approach can effectively alleviate the impact of anomalies without incurring prohibitive cost to the infrastructure. We focus on developing novel techniques for predicting, diagnosing, and correcting latent anomalies in shared computing infrastructures. The latent anomalies (e.g., performance degradation, resource hotspots, memory leak bugs) often do not have salient symptoms at the beginning, which make it hard to detect by human being. Those latent anomalies are often difficult to diagnose since their symptoms are often correlated with many reasons. However, it is highly important to detect and correct those latent anomalies since they often have prolonged impact to the system. We test our techniques on not only controllable virtual computing systems running in our lab but also on production-level infrastructures such as virtual computing lab (VCL) at NCSU and real world computing infrastructure data provided by our industrial partners at Google and IBM. We also develop metrics and models to evaluate the predictability of a wide range of system anomalies so as to build taxonomy of predictable system anomalies. We also develop new prediction and containment techniques to prevent root exploit attacks on edge-devices such as smart phones, which are the most serious attacks among all the security attacks and are hard to prevent using exiting techniques.

Scientific Barriers

Statistical learning and detailed data analysis have recently been shown to be promising for automatic system status analysis. Our work leverages statistical learning and signal

processing techniques to achieve online anomaly prediction. The major challenge includes how to achieve high prediction accuracy under dynamic computing environments and raise early enough alerts before anomaly happens. We have developed various online anomaly prediction techniques to achieve this goal. We developed prediction algorithms using both supervised and unsupervised learning techniques. The unsupervised learning approach allows us to achieve online anomaly prediction without requiring anomaly training data. Thus, our techniques can predict both previously known and unknown anomalies. We also developed context-aware anomaly prediction techniques that can achieve much higher prediction accuracy for dynamic systems than previous schemes. We recently extend our prediction algorithm that can consider not only system-level metrics (e.g., CPU, memory, disk usage) but also system calls. By analyzing system calls, our prediction algorithm can successfully predict all the existing root exploit attacks on the Android smart phones.

Prediction enables us to trigger timely preventions (e.g., migration, resource scaling, inserting delays in system calls) before the user perceives serious impact from the anomaly. We developed various online anomaly prevention techniques using live virtual machine (VM) migrations and elastic resource scaling. Our prediction system not only can raise advance alerts but also provide root cause inference to identify what might be the root cause of the system anomaly (e.g., CPU hog, memory leak, disk contention). We can then invoke proper prevention actions accordingly. Since prediction might raise false alarms, we also develop validation schemes to reverse incorrect preventions.

Prediction also enables us to perform in-situ anomaly diagnosis that can identify anomaly root causes onsite. The advantage is that we don't need to reproduce the anomaly-inducing environments, which are often extremely difficult. We are developing onsite anomaly path inference and various root cause localization techniques. We can first localize the faulty components among many distributed system components. We then localize root cause functions using system call analysis. The basic idea is to learn the system call sequence patterns produced by different functions using frequent episode mining and then use those system call sequence patterns as signatures to identify root cause functions. The advantage of our approach is that we don't require source code or any high-overhead online system instrumentations. We also develop onsite failure path inference without requiring source code.

Virtual machines provide opportunities for us to monitor and control various applications running inside the computing infrastructure. Our work leverages virtual machines to perform out-of-box monitoring and control. One challenge we have addressed in this project is to achieve scalable runtime monitoring, which can continuously track different virtual machine (VM) execution data (e.g., performance counters, resource metrics, system calls, inter-component invocations) to provide comprehensive knowledge for anomaly prediction and diagnosis. We developed adaptive sampling and online compression techniques to achieve light-weight monitoring.

Significance

The proposed research fundamentally advances knowledge and understanding in the interdisciplinary field of applying machine learning and dynamic system analysis to improve the resilience of complex computing infrastructures. Enhancing the resilience of large-scale computing infrastructures, which is well recognized by ARO as one of its key computing challenges in future battle spaces. As more and more critical Army missions depend on IT infrastructure, it has become imperative to guarantee continuous system operation despite software/hardware failures and malicious attacks. As rapid advances in computing hardware have led to dramatic improvement in computer performance, the issues of reliability, availability, and manageability are becoming the nominating bottlenecks in IT infrastructure maintenance. The proposed research advances existing science and technology through novel techniques in support of self-evolving system modeling, online anomaly prediction, onsite anomaly diagnosis, and anomaly preventions for large-scale distributed computing infrastructure. The proposed research explores new approaches with novel applications of machine learning, speculative execution, and dynamic system analysis on system profiling, anomaly prediction and diagnosis, and development of new scalable techniques and tools to achieve resilient distributed computing systems. We will develop and make available implemented techniques and collected data, which will let other researchers and practitioners build on our results.

Accomplishments
(feel free to use a bulleted list here)

Publications:

- "PerfScope: Practical Online Server Performance Bug Inference in Production Cloud Computing Infrastructures",
  Daniel Dean, Hiep Nguyen, Xiaohui Gu, Hui Zhang, Junghwan Rhee, Nipun Arora, Geoff Jiang
  Proc. of ACM Symposium on Cloud Computing (**SOCC**), Seattle, WA, November, 2014. (acceptance rate: 29/119 = 24%)
- "PerfCompass: Toward Runtime Performance Anomaly Fault Localization for Infrastructure-as-a-Service Clouds",
  Daniel Dean, Hiep Nguyen, Peipei Wang, Xiaohui Gu,
  Proc. of USENIX Workshop on Hot Topics in Cloud Computing (**HotCloud**), Philadelphia, PA, June, 2014. (acceptance rate: 22/72 = 30.5%)

- "Insight: In-situ Online Service Failure Path Inference in Production Computing Infrastructures",
  Hiep Nguyen, Daniel J. Dean, Kamal Kc, Xiaohui Gu
  Proc. of USENIX Annual Technical Conference (**USENIX ATC**), Philadelphia, PA, June, 2014. (acceptance rate: 36/241 = 14.9%)

- "PREC: Practical Root Exploit Containment for Android Devices", Tsung-Hsuan Ho, Daniel Dean, Xiaohui Gu, William Enck, Proc. of the ACM Conference on Data and

Application Security and Privacy (**CODASPY**), San Antonio, TX, March, 2014. (full paper, acceptance rate: 16%)

- "AGILE: elastic distributed resource scaling for Infrastructure-as-a-Service", Hiep Nguyen, Zhiming Shen, Xiaohui Gu, Sethuraman Subbiah, John Wilkes, Proc. of USENIX International Conference on Autonomic Computing (**ICAC**), San Jose, CA, June, 2013. (full paper, acceptance rate: 16/73 = 21%)
- "FChain: Toward Black-box Online Fault Localization for Cloud Systems", Hiep Nguyen, Zhiming Shen, Yongmin Tan, Xiaohui Gu, Proc. of IEEE International Conference on Distributed Computing Systems (**ICDCS**), Philadelphia, PA, July, 2013. (acceptance rate: 61/464 = 13%)
- "Scalable Distributed Service Integrity Attestation for Software-as-a-Service Clouds", Juan Du, Daniel Dean, Yongmin Tan, Xiaohui Gu, Ting Yu, IEEE Transactions on Parallel and Distributed Systems (**TPDS**), 2013.
- "UBL: Unsupervised Behavior Learning for Predicting Performance Anomalies in Virtualized Cloud Systems", Daniel Dean, Hiep Nguyen, Xiaohui Gu, Proc. of International Conference on Autonomic Computing (**ICAC**), San Jose, CA, September, 2012. (acceptance rate: 24%)
- "PREPARE: Predictive Performance Anomaly Prevention for Virtualized Cloud Systems", Yongmin Tan, Hiep Nguyen, Zhiming Shen, Xiaohui Gu, Chitra Venkatramani, Deepak Rajan, Proc. of International Conference on Distributed Computing Systems (**ICDCS**), Macau, China, June, 2012 (acceptance rate: 71/515=13.8%, *best paper award*).
- "Resilient Self-Compressive Monitoring for Large-Scale Hosting Infrastructures", Yongmin Tan, Vinay Venkatesh, Xiaohui Gu, IEEE Transactions on Parallel and Distributed Systems (**TPDS**), 2012.
- "Propagation-aware Anomaly Localization for Cloud Hosted Distributed Applications", Hiep Nguyen and Yongmin Tan and Xiaohui Gu, Proc. of ACM Workshop on Managing Large-Scale Systems via the Analysis of System Logs and the Application of Machine Learning Techniques (**SLAML**) in conjunction with **SOSP**, Cascais, Portugal, October, 2011.
- "ELT: Efficient Log-based Troubleshooting System for Cloud Computing Infrastructures", Kamal Kc, Xiaohui Gu, Proc. of IEEE International Symposium on Reliable Distributed Systems (**SRDS**), Madrid, Spain, October, 2011.
- "OLIC: OnLine Information Compression for Scalable Distributed System Monitoring", Yongmin Tan, Vinay Venkatesh, Xiaohui Gu, Proc. of ACM/IEEE International Workshop on Quality of Service (**IWQoS**), San Jose, CA, June, 2011.
- "Adaptive Runtime Anomaly Prediction for Dynamic Hosting Infrastructures", Yongmin Tan, Xiaohui Gu, Haixun Wang, ACM Symposium on Principles of Distributed Computing (**PODC**), Zurich, Switzerland, July, 2010. (Acceptance rate: 21%)
- "PRESS: PRedictive Elastic ReSource Scaling for Cloud Systems", Zhenhuan Gong, Xiaohui Gu, John Wilkes, IEEE International Conference on Network and Services Management (**CNSM**), Niagara Falls, Canada, October, 2010.(acceptance rate: 27/176 = 15%, ***Best Paper Award***)
- "On Predictability of System Anomalies in Real World", Yongmin Tan, Xiaohui Gu, Proc. of IEEE/ACM International Symposium on Modeling, Analysis and

Simulation of Computer and Telecommunication Systems (**MASCOTS**), Miami Beach, Florida, August, 2010. (Acceptance rate: 29%)

- "Self-Correlating Predictive Information Tracking for Large-Scale Production Systems", Ying Zhao, Yongmin Tan, Zhenhuan Gong, Xiaohui Gu, Mike Wamboldt, IEEE International Conference on Autonomic Computing and Communications (**ICAC**), Barcelona, Spain, June, 2009. (Acceptance rate: 15.6%)

Awards:

- Best paper awards, IEEE ICDCS, 1 out of 530 submissions, 2012.
- Best paper awards, IEEE CNSM, 1 out of 176 submissions, 2010.

Media coverage:

- featured highlight on NSF's official news site, Science 360,
- Communications of ACM,
- e! science news,
- WRAL techwire,
- ScienceDaily, etc.

Collaborations and Leveraged Funding

We have been collaborating with VCL administrators to apply our techniques on the VCL infrastructure. Most of our tools have been tested on the VCL. We have been working with researchers at IBM and Google during this project. Our current leveraged funding include:

- "CAREER: Enabling Robust Virtualized Hosting Infrastructures via Coordinated Learning, Recovery, and Diagnosis", NSF, $450K, 1/1/2012-12/31/2016, Sole PI.
- "Deepening the Understanding of Least Privilege Through Automatic Partitioning of Hybrid Programs", NSA Science of Security Lablet, $522K, 1/1/2012-12/31/2014, Co-PI, PI: William Enck.
- "Online Performance Anomaly Diagnosis for Cloud Computing Infrastructures", IBM Faculty Award, $15K, 9/1/2011-8/31/2012, Sole PI.
- "CSR:Small: Online System Anomaly Prediction and Diagnosis for Large-Scale Hosting Infrastructures", NSF, $405,000, 08/15/2009 – 08/14/2012, Sole PI.

Conclusions

We have successfully integrated our online anomaly prediction, anomaly root cause inference, and anomaly prevention components into a complete automatic anomaly prevention framework. Our system can automatically steer the system away from anomalies caused by various software bugs, resource contentions, or malicious attacks.

Technology Transfer

We have deployed and tested most of our techniques on the virtual computing lab (VCL) at North Carolina State University. We also tested our system on a cloud computing testbed (HGCC cluster) in our lab that consists of 15 blade nodes. Recently, we have filed provisional patent and several software licenses on our technologies. Several big companies including Google have indicated their interests in licensing our technologies.

Future Plans

We will extensively test the current framework with various real world applications and malwares. We hope to provide a detailed study on which kind of malicious attacks can be captured by our framework. We will continue to develop more robust and practical online anomaly/malware prediction techniques to capture malicious activities. We will also develop malicious software sandboxing techniques that allow us to capture malicious activities without compromising the protected system.

Objective

Larg-scale distributed computing infrastructures have become important platforms for many critical real-world systems such as cloud computing, big data processing, and intelligence analysis. However, due to its inherent complexity and sharing nature, shared computing infrastructures are inevitably prone to various system anomalies caused by software bugs, hardware failures, and resource contentions. The situation exacerbates if the system is also exposed to malicious attacks. Moreover, although some anomaly symptoms such as machine crash are easy to detect, many other anomalies (e.g., performance degradation, processing bottlenecks, memory leak bugs) are hard to detect and diagnosis, which often have latent impact to the system. The objective of this project is to develop automatic 24x7 anomaly management to enhance the resilience of large-scale shared computing infrastructures.

Approach

In this project, we propose to develop a new predictive anomaly management approach that can raise advance anomaly alerts to trigger just-in-time anomaly diagnosis while the system approaches the anomaly state, and perform informed anomaly correction based on the runtime diagnosis results before the system is seriously affected by the anomaly. Thus, our approach can effectively alleviate the impact of anomalies without incurring prohibitive cost to the infrastructure. We focus on developing novel techniques for predicting, diagnosing,

and correcting latent anomalies in shared computing infrastructures. The latent anomalies (e.g., performance degradation, resource hotspots, memory leak bugs) often do not have salient symptoms at the beginning, which make it hard to detect by human being. Those latent anomalies are often difficult to diagnose since their symptoms are often correlated with many reasons. However, it is highly important to detect and correct those latent anomalies since they often have prolonged impact to the system. We test our techniques on not only controllable virtual computing systems running in our lab but also on production-level infrastructures such as virtual computing lab (VCL) at NCSU and real world computing infrastructure data provided by our industrial partners at Google and IBM. We also develop metrics and models to evaluate the predictability of a wide range of system anomalies so as to build taxonomy of predictable system anomalies. We also develop new prediction and containment techniques to prevent root exploit attacks on edge-devices such as smart phones, which are the most serious attacks among all the security attacks and are hard to prevent using exiting techniques.

Scientific Barriers

Statistical learning and detailed data analysis have recently been shown to be promising for automatic system status analysis. Our work leverages statistical learning and signal processing techniques to achieve online anomaly prediction. The major challenge includes how to achieve high prediction accuracy under dynamic computing environments and raise early enough alerts before anomaly happens. We have developed various online anomaly prediction techniques to achieve this goal. We developed prediction algorithms using both supervised and unsupervised learning techniques. The unsupervised learning approach allows us to achieve online anomaly prediction without requiring anomaly training data. Thus, our techniques can predict both previously known and unknown anomalies. We also developed context-aware anomaly prediction techniques that can achieve much higher prediction accuracy for dynamic systems than previous schemes. We recently extend our prediction algorithm that can consider not only system-level metrics (e.g., CPU, memory, disk usage) but also system calls. By analyzing system calls, our prediction algorithm can successfully predict all the existing root exploit attacks on the Android smart phones.

Prediction enables us to trigger timely preventions (e.g., migration, resource scaling, inserting delays in system calls) before the user perceives serious impact from the anomaly. We developed various online anomaly prevention techniques using live virtual machine (VM) migrations and elastic resource scaling. Our prediction system not only can raise advance alerts but also provide root cause inference to identify what might be the root cause of the system anomaly (e.g., CPU hog, memory leak, disk contention). We can then invoke proper prevention actions accordingly. Since prediction might raise false alarms, we also develop validation schemes to reverse incorrect preventions.

Prediction also enables us to perform in-situ anomaly diagnosis that can identify anomaly root causes onsite. The advantage is that we don’t need to reproduce the anomaly-inducing environments, which are often extremely difficult. We are developing onsite anomaly path inference and various root cause localization techniques. We can first localize the faulty components among many distributed system components. We then localize root cause

functions using system call analysis. The basic idea is to learn the system call sequence patterns produced by different functions using frequent episode mining and then use those system call sequence patterns as signatures to identify root cause functions. The advantage of our approach is that we don't require source code or any high-overhead online system instrumentations. We also develop onsite failure path inference without requiring source code.

Virtual machines provide opportunities for us to monitor and control various applications running inside the computing infrastructure. Our work leverages virtual machines to perform out-of-box monitoring and control. One challenge we have addressed in this project is to achieve scalable runtime monitoring, which can continuously track different virtual machine (VM) execution data (e.g., performance counters, resource metrics, system calls, inter-component invocations) to provide comprehensive knowledge for anomaly prediction and diagnosis. We developed adaptive sampling and online compression techniques to achieve light-weight monitoring.

Significance

The proposed research fundamentally advances knowledge and understanding in the interdisciplinary field of applying machine learning and dynamic system analysis to improve the resilience of complex computing infrastructures. Enhancing the resilience of large-scale computing infrastructures, which is well recognized by ARO as one of its key computing challenges in future battle spaces. As more and more critical Army missions depend on IT infrastructure, it has become imperative to guarantee continuous system operation despite software/hardware failures and malicious attacks. As rapid advances in computing hardware have led to dramatic improvement in computer performance, the issues of reliability, availability, and manageability are becoming the nominating bottlenecks in IT infrastructure maintenance. The proposed research advances existing science and technology through novel techniques in support of self-evolving system modeling, online anomaly prediction, onsite anomaly diagnosis, and anomaly preventions for large-scale distributed computing infrastructure. The proposed research explores new approaches with novel applications of machine learning, speculative execution, and dynamic system analysis on system profiling, anomaly prediction and diagnosis, and development of new scalable techniques and tools to achieve resilient distributed computing systems. We will develop and make available implemented techniques and collected data, which will let other researchers and practitioners build on our results.

Accomplishments

(feel free to use a bulleted list here)

Publications:

- "PerfScope: Practical Online Server Performance Bug Inference in Production Cloud Computing Infrastructures", Daniel Dean, Hiep Nguyen, Xiaohui Gu, Hui Zhang, Junghwan Rhee, Nipun Arora, Geoff Jiang, Proc. of ACM Symposium on Cloud Computing (SOCC), Seattle, WA, November, 2014. (acceptance rate: 29/119 = 24%)

- "PerfCompass: Toward Runtime Performance Anomaly Fault Localization for Infrastructure-as-a-Service Clouds", Daniel Dean, Hiep Nguyen, Peipei Wang, Xiaohui Gu, Proc. of USENIX Workshop on Hot Topics in Cloud Computing (HotCloud), Philadelphia, PA, June, 2014. (acceptance rate: 22/72 = 30.5%)
- "Insight: In-situ Online Service Failure Path Inference in Production Computing Infrastructures", Hiep Nguyen, Daniel J. Dean, Kamal Kc, Xiaohui Gu
Proc. of USENIX Annual Technical Conference (USENIX ATC), Philadelphia, PA, June, 2014. (acceptance rate: 36/241 = 14.9%)
- "PREC: Practical Root Exploit Containment for Android Devices", Tsung-Hsuan Ho, Daniel Dean, Xiaohui Gu, William Enck, Proc. of the ACM Conference on Data and Application Security and Privacy (CODASPY), San Antonio, TX, March, 2014. (full paper, acceptance rate: 16%)
- "AGILE: elastic distributed resource scaling for Infrastructure-as-a-Service", Hiep Nguyen, Zhiming Shen, Xiaohui Gu, Sethuraman Subbiah, John Wilkes, Proc. of USENIX International Conference on Autonomic Computing (ICAC), San Jose, CA, June, 2013. (full paper, acceptance rate: 16/73 = 21%)
- "FChain: Toward Black-box Online Fault Localization for Cloud Systems", Hiep Nguyen, Zhiming Shen, Yongmin Tan, Xiaohui Gu, Proc. of IEEE International Conference on Distributed Computing Systems (ICDCS), Philadelphia, PA, July, 2013. (acceptance rate: 61/464 = 13%)
- "Scalable Distributed Service Integrity Attestation for Software-as-a-Service Clouds", Juan Du, Daniel Dean, Yongmin Tan, Xiaohui Gu, Ting Yu, IEEE Transactions on Parallel and Distributed Systems (TPDS), 2013.
- "UBL: Unsupervised Behavior Learning for Predicting Performance Anomalies in Virtualized Cloud Systems", Daniel Dean, Hiep Nguyen, Xiaohui Gu, Proc. of International Conference on Autonomic Computing (ICAC), San Jose, CA, September, 2012. (acceptance rate: 24%)
- "PREPARE: Predictive Performance Anomaly Prevention for Virtualized Cloud Systems", Yongmin Tan, Hiep Nguyen, Zhiming Shen, Xiaohui Gu, Chitra Venkatramani, Deepak Rajan, Proc. of International Conference on Distributed Computing Systems (ICDCS), Macau, China, June, 2012 (acceptance rate: 71/515=13.8%, best paper award).
- "Resilient Self-Compressive Monitoring for Large-Scale Hosting Infrastructures", Yongmin Tan, Vinay Venkatesh, Xiaohui Gu, IEEE Transactions on Parallel and Distributed Systems (TPDS), 2012.
- "Propagation-aware Anomaly Localization for Cloud Hosted Distributed Applications",
Hiep Nguyen and Yongmin Tan and Xiaohui Gu,
Proc. of ACM Workshop on Managing Large-Scale Systems via the Analysis of System Logs and the Application of Machine Learning Techniques (SLAML) in conjunction with SOSP, Cascais, Portugal, October, 2011.
- "ELT: Efficient Log-based Troubleshooting System for Cloud Computing Infrastructures",
Kamal Kc, Xiaohui Gu,
Proc. of IEEE International Symposium on Reliable Distributed Systems (SRDS), Madrid, Spain, October, 2011.

- "OLIC: OnLine Information Compression for Scalable Distributed System Monitoring", Yongmin Tan, Vinay Venkatesh, Xiaohui Gu, Proc. of ACM/IEEE International Workshop on Quality of Service (IWQoS), San Jose, CA, June, 2011.
- "Adaptive Runtime Anomaly Prediction for Dynamic Hosting Infrastructures", Yongmin Tan, Xiaohui Gu, Haixun Wang,
ACM Symposium on Principles of Distributed Computing (PODC), Zurich, Switzerland, July, 2010. (Acceptance rate: 21%)
- "PRESS: PRedictive Elastic ReSource Scaling for Cloud Systems", Zhenhuan Gong, Xiaohui Gu, John Wilkes, IEEE International Conference on Network and Services Management (CNSM), Niagara Falls, Canada, October, 2010.(acceptance rate: 27/176 = 15%, Best Paper Award)
- "On Predictability of System Anomalies in Real World", Yongmin Tan, Xiaohui Gu,
Proc. of IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS), Miami Beach, Florida, August, 2010. (Acceptance rate: 29%)
- "Self-Correlating Predictive Information Tracking for Large-Scale Production Systems", Ying Zhao, Yongmin Tan, Zhenhuan Gong, Xiaohui Gu, Mike Wamboldt,
IEEE International Conference on Autonomic Computing and Communications (ICAC), Barcelona, Spain, June, 2009. (Acceptance rate: 15.6%)

Awards:
- Best paper awards, IEEE ICDCS, 1 out of 530 submissions, 2012.
- Best paper awards, IEEE CNSM, 1 out of 176 submissions, 2010.

Media coverage:
- featured highlight on NSF's official news site, Science 360,
- Communications of ACM,
- e! science news,
- WRAL techwire,
- ScienceDaily, etc.

Collaborations and Leveraged Funding

We have been collaborating with VCL administrators to apply our techniques on the VCL infrastructure. Most of our tools have been tested on the VCL. We have been working with researchers at IBM and Google during this project. Our current leveraged funding include:

- "CAREER: Enabling Robust Virtualized Hosting Infrastructures via Coordinated Learning, Recovery, and Diagnosis", NSF, $450K, 1/1/2012-12/31/2016, Sole PI.
- "Deepening the Understanding of Least Privilege Through Automatic Partitioning of Hybrid Programs", NSA Science of Security Lablet, $522K, 1/1/2012-12/31/2014, Co-PI, PI: William Enck.
- "Online Performance Anomaly Diagnosis for Cloud Computing Infrastructures", IBM Faculty Award, $15K, 9/1/2011-8/31/2012, Sole PI.

- "CSR:Small: Online System Anomaly Prediction and Diagnosis for Large-Scale Hosting Infrastructures",
NSF, $405,000, 08/15/2009 – 08/14/2012,
Sole PI.

Conclusions

We have successfully integrated our online anomaly prediction, anomaly root cause inference, and anomaly prevention components into a complete automatic anomaly prevention framework. Our system can automatically steer the system away from anomalies caused by various software bugs, resource contentions, or malicious attacks.

Technology Transfer

NCSU filed a patent application on our unsupervised anomaly prediction scheme and Google has purchased an evaluation license for our software.